Project: Machine Learning

https://github.com/santhinagalla/Machine-Learning/tree/main/Supervised%20Learning/Linear%20Regression

1.What is Linear Regression?

Linear regression is a statistical modeling technique used to describe a continuous response variable as a function of one or more predictor variables. It can help you understand and predict the behavior of complex systems or analyze experimental, financial, and biological data.

Linear regression techniques are used to create a linear model. The model describes the relationship between a dependent variable y (also called the response) as a function of one or more independent variables Xi (called the predictors). The general equation for a linear regression model is:

$$Y=\beta 0+\sum \beta iXi+\epsilon i$$

where β represents linear parameter estimates to be computed and ϵ represents the error terms.

Regression Formula: (another formula produces the same result) :

Regression Equation(y) = a + bx

Slope(b) = (NΣXY - (ΣX) (ΣY)) / (NΣX2 - (ΣX)2)

Intercept(a) = (ΣY - b(ΣX)) / N

Where:

x and y are the variables.

b = The slope of the regression line

a = The intercept point of the regression line and the y axis.

N = Number of values or elements

X = First Score

Y = Second Score

ΣXY = Sum of the product of first and Second Scores

ΣX = Sum of First Scores

ΣY = Sum of Second Scores

ΣX2 = Sum of square First Scores

To find the Simple/Linear Regression of below data -

X Values	Y Values
60	3.1
61	3.6
62	3.8
63	4
65	4.1

Steps to find regression equation, we will first find slope, intercept and use it to form regression equation.

- Count the number of values. In our example N = 5
- Find X * Y, X^2
- Find ΣX, ΣY, ΣXY, ΣX^2.
- Substitute in the above slope formula given Slope(b) = (NΣXY - (ΣX) (ΣY)) / (NΣX2 - $(\Sigma X)^2$)
- Intercept(a) = (ΣY - b(ΣX)) / N
- Then substitute Intercept(a) and Slope(b) in regression equation formula **Regression Equation(y) = a + bx**

Java Program to Implement Linear Regression:

Steps to follow -

- Collect the data: Read the data from the file "input.text"
- Create the model: Find "a" and "b" of a "Linear Regression Equation

  (y) = a + bx" based on the content of "input.txt".

- Prediction: If x=64, predict the y value.

Java Program

```
import java.io.*;

import java.io.FileReader;

import java.io.BufferedReader;
```

```
public class Linear_Regression {
   double b =0, a=0;
                //Substitute calculated values in Slope formula
            "Slope(b) = (NΣXY - (ΣX)(ΣY)) / (NΣX2 - (ΣX)2)"
   public double slope(int N,double xy,double x,double y,double xx){
      b = ((N * xy) - (x * y))/((N * xx) - (x * x));
      return b;
   }
       //Substitute calculated values in intercept formula
                "Intercept(a) = (ΣY - b(ΣX)) /N"
   public double intercept(int N,double x,double y){
      a = (y - b * x)/N;
      return a;
   }
   //Then substitute Intercept(a) and Slope(b) in regression equation formula
        "Regression Equation(y) = a + bx"
   public double calculateRegression(double variable_X){
      double predict_Y = a + b * variable_X;
      return predict_Y;
   }
  public static void main(String args[]) throws IOException {
                int noOfLines= 0; //Count the number of values
                  // Read the data from the file "input.text"
 BufferedReader bufReader = new BufferedReader(new FileReader("input.txt"));
      double x_sum=0, y_sum = 0, xy_sum =0,x2_sum=0;
      String line = bufReader.readLine();
```

```
while (line != null) {
    String[] fields = line.split("\\s+");
    double x = Double.parseDouble(fields[0]);
    double y = Double.parseDouble(fields[1]);
    double xy = x*y; //Find X * Y, X2
    double x2 = x*x;
    x_sum +=x;
    y_sum +=y;
    xy_sum += xy;
    x2_sum += x2;
    line = bufReader.readLine();
    noOfLines++;
}
bufReader.close();
//Find ΣX, ΣY, ΣXY, ΣX2.
System.out.println("Sum of X  = " + x_sum);
System.out.println("Sum of Y  = " + y_sum);
System.out.println("Sum of XY  = "+ xy_sum);
System.out.println("Sum of X^2 = " + x2_sum);

Linear_Regression linear = new Linear_Regression();
System.out.println("Slope = " + linear.slope(noOfLines, xy_sum, x_sum, y_sum,x2_sum)); //Display Slope Value
System.out.println("Intercept = " + linear.intercept(noOfLines, x_sum, y_sum)); //Display Intercept Value
```

//Suppose if we want to know the approximate y value for the variable x = 64. Then we can substitute the value in the "calculateRegression" equation.

```
    double predict= linear.calculateRegression(64);
   System.out.print("Using Regression Equation if x= 64, predict the Y value = ");
    System.out.printf("%.2f", predict);
      }
  }
```

Output -

Sum of X = 311.0

Sum of Y = 18.6

Sum of XY = 1159.7

Sum of X^2 = 19359.0

Slope = 0.18783783783783292

Intercept = -7.963513513513208

Using Regression Equation if x= 64, predict the Y value = 4.06